DOCUMENT RESUME

ED 464 196 UD 034 991

AUTHOR	Gorard, Stephen; Taylor, Chris
TITLE	A Preliminary Consideration of the Impact of Market Forces on Educational Standards. Measuring Markets: The Case of the ERA 1988. Working Paper.
INSTITUTION	Cardiff Univ. (Wales). School of Social Sciences.
SPONS AGENCY	Economic and Social Research Council, Lancaster (England).
REPORT NO	WP-38
ISBN	ISBN-1-87-2330-355
PUB DATE	2001-00-00
NOTE	23p.
CONTRACT	R000238031
AVAILABLE FROM	For full text: http://www.cardiff.ac.uk/socsi/markets.
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Academic Achievement; *Academic Standards; Competition; Educational Research; Foreign Countries; *Free Enterprise System; Poverty; Research Methodology; *School Choice; Secondary Education
IDENTIFIERS	England; Wales

ABSTRACT

This paper considers the possible impact of market forces on educational achievement in English and Welsh secondary schools. A major argument made by market advocates in favor of extending programs of school choice was that this would drive up standards. However, despite 12 years of relevant experience in the United Kingdom, it remains very difficult to test this claim. This paper examines practical difficulties in conducting research, then presents three possible models for considering changes in academic standards over time. It discusses what an appropriate control group would be when assessing the impact of choice and how researchers can otherwise control for the impact of student background. The results are inconclusive, and possibly even contradictory. The measures, such as GCSE and A levels, extending back to 1988 and beyond, have increased in prevalence. In terms of these measures, students from state-funded schools have reduced the gap between their achievement and that of students from fee-paying institutions. However, it is not clear that these developments are market related. There is also no evidence that these improvements indicate any weakness in the strong link between the socioeconomic background of students and their school outcomes. (Contains 57 references, 3 figures, and 1 table.) (SM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 464 196

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

MEASURING MARKETS: THE CASE OF THE ERA 1988

www.cardiff.ac.uk/socsi/markets

A Cardiff University School of Social Sciences ESRC-funded Project

WORKING PAPER 38

A preliminary consideration of the impact of market forces on educational standards

Stephen Gorard and Chris Taylor

2001

Cardiff University School of Social Sciences
Glamorgan Building
King Edward VII Avenue
Cardiff
CF10 3WT

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

C. Taylor
Cardiff Univ. School of Social Sciences

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Email: gorard@cardiff.ac.uk

ISBN 1 872330 355

UD 034 991

BEST COPY AVAILABLE

A preliminary consideration of the impact of market forces on educational standards

Stephen Gorard and Chris Taylor

Cardiff University School of Social Sciences
King Edward VII Avenue
Cardiff, CF10 3WT
02920-875113
email: gorard@cardiff.ac.uk

Abstract

This paper considers the possible impact of market forces on educational attainment in secondary schools in England and Wales. One of the main arguments made by market advocates in favour of extending programmes of school choice was that this would drive up standards. However, despite twelve years of relevant experience in the UK it remains very difficult to test this claim. The paper examines some practical difficulties before presenting three possible models for considering changes in educational standards over time. The results are inconclusive, possibly even contradictory. The measures, such as GCSE and A levels, extending back to 1988 and beyond have clearly increased in prevalence. In terms of these measures, students from state-funded education have also reduced the 'gap' relative to those from fee-paying institutions. However, it is not clear that either of these developments is market-related. In addition, there is no evidence yet that these improvements indicate any breakage in the strong link between the socio-economic background of students and their school outcomes.

Introduction - choice policy and claims

The various justifications, and consequent policies, for extending the notion of parental choice of secondary schools in the 1980s and beyond have been described in detail elsewhere (e.g. Gorard 1999a). The writing by advocates of choice programmes in education falls quite neatly into three arguments (e.g. Friedman and Friedman 1980, see also Witte 1990). First: there is the libertarian notion of choice for its own sake (Erickson 1989). We all appreciate choice as consumers in some areas, so why not others? This approach is apparently justified by the popularity of school choice programmes expressed in opinion polls, and in the increasing participation of many sections of society after policies have been introduced. Second: there is the argument for equity (Cookson 1994). Choice of school extends a privilege to all that was previously available only to those able to afford houses in desirable suburban catchment areas, or to send their child to a fee-paying school. This approach is apparently justified by the declining socio-economic segregation in the school systems of England and Wales from 1988.

The third argument, which is perhaps the most important for choice advocates, is that market forces will drive up educational standards (Chubb and Moe 1990). Successful schools will be popular, and weaker schools will be unpopular, progressively losing their *per capita* funding until they either improve or close. Over time therefore the general standard of schools will be higher. However, although critics of market forces have themselves been criticised for not providing any evidence for the effectiveness of state-funded service monopolies (Friedman and Friedman 1980) there is also no contrary evidence yet of the effectiveness of choice. As Fuller et al. (1996) put it - will choice programmes create more effective forms of schools, and will standards rise? Given that the Education Reform Act 1988, the pivotal legislation school choice in England and Wales, is now 12 years old what evidence is there that standards have risen? Put another way - what evidence could there be?

Difficulties

The first practical difficulty to be faced lies in deciding precisely what is meant by 'standards' in schools and how these are measured. We need an explicit form which is comparable over time. In a

compulsory system we cannot use the general popularity of education as an indicator of its success. Many of the potential outputs of education are either so long-term (a preparation for later life) or so nebulous (educating the whole person) that they are unusable in this way. These 'softer', wider and longer-term outcomes of education are clearly important, even though they are not susceptible to rigorous analysis of the claims of market advocates. There remain two obvious measures that could be used - school examination outcomes, and financial efficiency. While the implications for the second can be followed through the work of Hardman and Levacic 1997 and West et al. 2000, this paper focuses on the first of these issues. Have schools produced higher levels of qualifications as a result of 12 years of parental choice?

Comparability

The difficulties of making comparisons between qualifications over time and place are well-documented (see Gorard 2000a). Differences between notionally equivalent qualifications, changes in content and methods of assessment over time, and, above all, the basic unreliability of assessment procedures make any claims of comparability suspect even in terms of the narrow educational outcomes measured by qualifications. Using the 'catch-all' definition of standards suggested by Baird et al. (2000), it is currently very difficult to answer the question 'Have exam standards fallen over time?'. If all other variables are held constant then do similar - since we cannot re-test the same candidate - candidates get the same result on two examinations? We have, up to now, had insufficient data to address this question.

Britain is peculiar among OECD members in using different regional authorities (local examination boards) to examine what are meant to be national assessments at 16+ and 18+ (Noah and Eckstein 1992). This raises an issue of whether the same qualification is equivalent between these boards in terms of difficulty. It is already clear that even qualifications with the same name (e.g. GCSE History) are not equivalent in terms of subject content as each board sets its own syllabus. Nor are they equivalent in the form of assessment, or the weighting between components such as coursework and multiple-choice. Nor is there any evidence that the different subjects added together to form aggregate benchmarks are equivalent in difficulty to each other, yet the standard GCSE benchmark gives the same value to an A* in Music, a B in Physics, and a C grade in Sociology. Nor is there evidence that qualifications with the same name are equally difficult from year to year. In fact, comparability can be considered between boards in any subject, the years in a

subject/board combination, the subjects in one board, and the alternative syllabuses in any board and subject. All of these are very difficult to determine, especially as exams are neither accurate nor particularly reliable in what they measure (Nuttall 1979). Pencil-and-paper tests have little generalisable validity, and their link to other measures such as occupational competence is also generally very small.

Gorard et al. (1999a) list many problems relating to the comparability of forms of assessment, and new stories of the unreliability of examinations appear in the media almost every week (e.g. Mansell 2000). As many as 12,500 of the 17,600 candidates taking the 11 plus in Northern Ireland could have been given wrongly classified grades (McGill 2000). The papers were marked out of 150. The standard error in the marking was 20 marks, but the top and bottom grades were separated by only 18 marks (i.e. the total range of achieved marks was less than the standard error of the marking). In the same year, one in seven of the primary schools in England asked for remarks of their English papers at Key Stage Two (Cassidy et al. 2000). As a result, 4,000 papers were upgraded and many markers were made redundant as being too unreliable. At GCSE, three years of results tables from Manchester LEA are now considered unsound, and not comparable with other LEAs, since the schools had removed students with persistent unauthorised absence from their rolls. As many as 7% of the students are therefore missing from the league tables, presumably leading to considerable inflation of the scores.

There are additional problems, other than changes in examinations and curricula, in comparing the results of schools over time. Although there are some indications that literacy standards in primary schools are improving, these could simply be due to increasingly lenient marking according to a number of headteachers - leading to wildly unrealistic grades in some cases (Hackett and Kelly 2000). In the 21 years since the Warnock Report nearly a quarter of special schools have disappeared with a similar decline in the number of full-time pupils attending them (Howson 1999). This decline runs alongside a large growth in indicators relating to special needs. Thus each mainstream school now contains a higher proportion of SEN students than it did in previous years, with a consequent impact on raw-score indicators of school 'performance'.

Where progress is measured, not in raw-score terms, but in terms of improvement from one Key Stage at school to the next then the differences between subjects and levels can make the earlier

'performance' a poor predictor of later performance (Robinson 1997), worse as a predictor than levels of poverty for example (Schagen and Morrison 1998).

Background

The difficulties of deciding which background factors, if any, to control for when making value-added comparisons between students, schools, or assessments are considerable (Willms 1992). There are clear systematic differences in attainment between social groups identified by language, gender, class, ethnicity and so on. Yet many of these identifiable factors interact in such a way that any one of them can be shown to be 'redundant' as a background factor, appearing instead as a proxy for some other combination of factors.

For example, at Key Stage 3 and 4 there are clear differences in attainment by ethnic group, but once other background factors, such as class, are accounted for then ethnicity has little direct effect (Haque and Bell 2000). Similarly with pupil mobility (turnover within years between schools). Standards in some of poorest regions appear to be affected by high pupil turnover (Dobson and Henthorne 1999). One school lost more than 40% of pupils in one year, but unlike family poverty and language requirements the current funding arrangements do not recognise this problem. Perhaps rightly so, for once other indicators of disadvantage and pupil prior attainment are used, then the direct effect of mobility disappears (Strand 2000). Once prior attainment is accounted for at student and institution level, there is no difference between the effectiveness of different school type such as grammar and comprehensive (Yang and Woodhouse 2000).

Changes over time

Even if one assumes that test and GCSE scores are a real measure of standards with equivalencies that can be compared across place and time, there are still complexities in creating an appropriate 'index' for comparisons. The TES (1999) reported that educational action zones (EAZs) have been successful in raising standards faster than their neighbours, in terms of test and GCSE scores. The report showed no concern about issues of reliability, but even granted this its conclusions are invalid, being based on what we have termed 'the politicians error' of assuming percentage points to be at an equal-interval level of measurement even where the figures from which they are drawn are themselves changing (Gorard 1999b, 1999c). The evidence presented is that, in raw-score terms,

the growth of examination scores in EAZs is greater than at local schools and the national average. This may well be so, but cannot be substantiated simply in terms of percentage point differences.

The limited evidence available so far on changes over time in school standards, expressed in examination terms, is far from clear. Evidence from experiments with school choice in the USA tends to suggest that choice is associated with small achievement gains (Powers and Cookson 1999, Jeynes 2000). For example, there is a correlation between school choice and improvements in reading and mathematics scores, just as there was in the private schools in the Coleman et al. (1982) study. Choice may be especially effective for ethnic minority groups, who might need the most help and also show the greatest gain as a result (Jeynes 2000).

However, as suggested above and demonstrated below, the barriers to obtaining a firm answer to questions about changes in standards are considerable. School effects, small as they are in relation to socio-economic determinants, show little stability from year-to-year (Yang and Woodhouse 2000). If choice reforms are accompanied by other changes in an educational system it becomes difficult to isolate the actual cause of academic improvement. In the USA, most research is based on surveys (Powers and Cookson 1999). Some of these simply show improvements as perceived (or reported) changes over time, and some do not even attempt to disentangle the impact of family background. The ensuing methodological debates there have led to two key prior questions in this area of research.

- What is an appropriate control group when assessing the impact of choice?
- How can we otherwise control for the impact of student background?

This paper addresses both of these questions.

Raw-score improvements

The first, and perhaps the simplest, way of expressing changes over time is to consider the prevalence of particular qualifications over successive age cohorts. The data presented below come from the DfEE (1998), and represent the GCSE/GCE/CSE results of all school-leavers in England from 1974/75 to 1997/98. There have been some important changes in the collection of these figures over time, in the definition of the relevant age cohort and, of course, in the nature of the

qualifications, but these data currently represent the best estimates available. The analysis focuses on the proportion of the relevant school-leaver cohort obtaining one or more GCSE grade G and above (or a CSE), and those obtaining five or more GCSEs grade A*-C (or an O level, or CSE grade 1).

If the equivalencies between various qualifications and modes of examination are not valid, then their comparisons over time are impossible and the market thesis will have to remain untested (and therefore opinions on it either way will be in the nature of superstitions). If, on the other hand, the equivalencies are deemed reasonable, then it is clear that the introduction of market forces has had no beneficial impact on school standards at low levels of attainment. Figure 1 shows the proportion of the age cohort obtaining any nationally recognised 16+ qualification for each year 1975-1998. These data show a clear but lessening rise in attainment until 1984, a plateau until 1988, a short rise in 1989, and a further plateau until 1998 (we are still seeking explanations for the dip in 1992). The rise in 1989 is unlikely to be attributable to the impact of the ERA 1988, which can only have come into force in 1989 at the earliest. The GCSE was introduced in 1986, so that the first examination after a two-year course took place in 1988. If the first cohort recruited via school choice was in 1989, they would not generally have taken their GCSEs in 1993. The attainment figure for 1993 is lower than in 1991. Even if closer analysis suggests a slight improvement in the annual qualification rate after 1988, it is clear that the improvement rate before 1988 was considerably greater. On the other hand it would be almost as difficult to argue that market forces have had a deleterious effect on attainment using these figures since, *ceteris paribus*, one would expect the rate of improvement to decline as the limit of 100% nears (similar to the notion of 'regression towards the mean', see Gorard 2000b). While 1989 was the year of greatest growth a number of related changes took place in that same period in addition to the implementation of policy extending parental choice (see below).

Anyway, it could be argued that the use of 'league tables' of results focusing on achievement at GCSE C grade and above, while not necessarily entailed in the programme of choice, means that it is at that level that any improvement could be expected. The trend for this level of attainment (C) is quite different to that for grade G. Figure 2 shows the percentage of students attaining the benchmark of five 'good' GCSEs or equivalent over time. Before 1988 there is little annual growth in this indicator. After 1988 the indicator is significantly higher for each successive year. Is this evidence of improvement due to choice?

Figure 1 - Percentage attaining 1+ GCSE A*-G equivalent

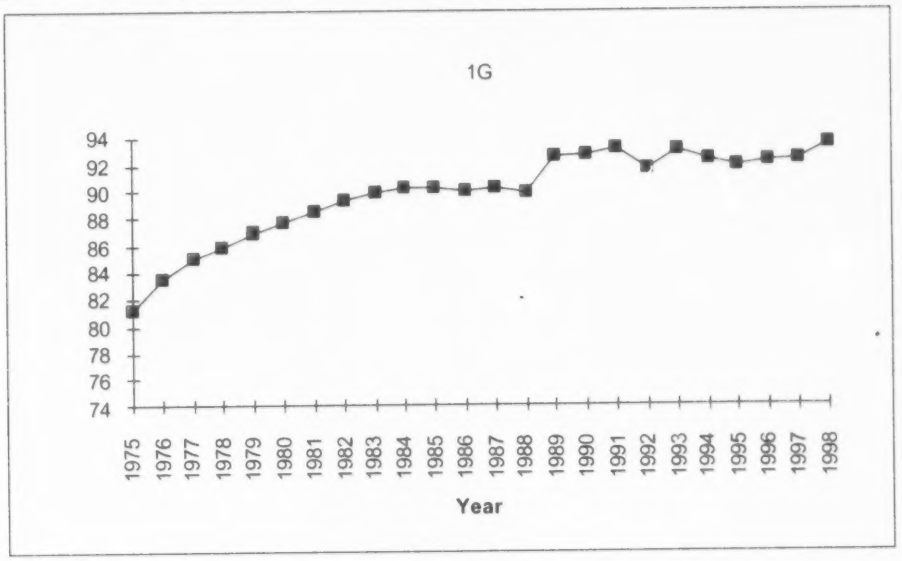
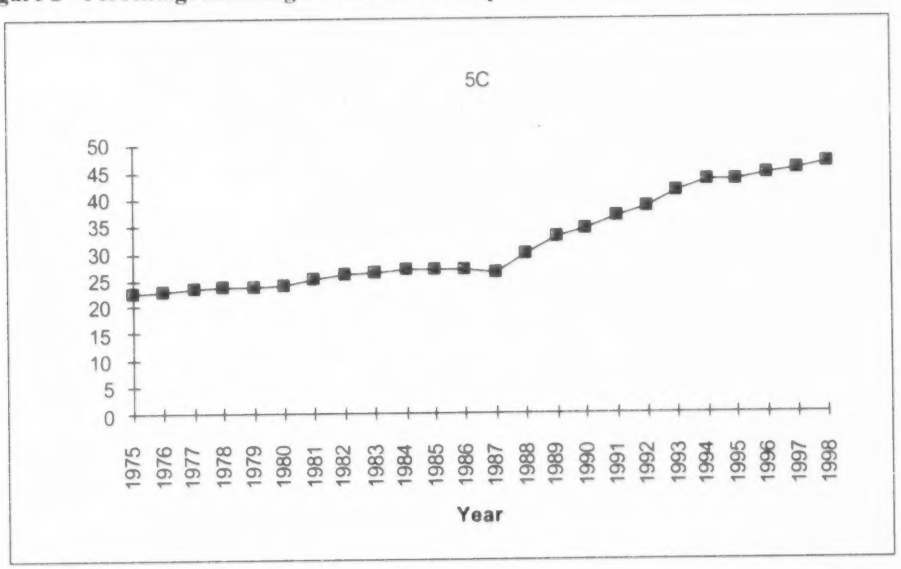


Figure 2 - Percentage attaining 5+ GCSE A*-C equivalent



The answer is likely to be 'no'. Again the growth appears too early to be the result of a 1988 policy change. In addition, this period involved so many other changes of direct relevance to the system of

assessment that it is difficult to unpick the threads of each. The introduction of the GCSE in 1986 heralded several major changes in the forms of assessment, most notably an increase in coursework at the expense of terminal examinations. 1987 was also the year in which strict norm-referencing - allocating grades in proportion to the entry cohort - was abolished in favour of increased criterion referencing - allocating grades in terms of skills and competencies. In addition, the publication of the results for the entire 15-year-old cohort replaced the previous School Leavers Survey (which had included a few results from children of other age-groups) at this time, and formed the basis for new school performance tables. All of these concurrent changes were linked to the largest ever annual increase in the reported proportion of those reaching the GCSE (or O level equivalent) benchmark in 1988, and the second largest in 1989.

Norm-referencing is unfair in a system in which not all subjects are taken by the same candidates. Problems of selection and self-selection bias mean that giving the same distribution of grades in Ancient Greek as in Media Studies for example cannot be justified. Criterion referencing is an example of what Baird et al. (2000) call a 'sociological perspective' on comparability, allowing expert judges to decide on comparisons since standards only exist as social constructs anyway. Where norm-referencing is not used, and scores are allowed to increase annually there is thus a danger of producing 'counterfeit excellence' (Zirkel 1999). In the USA high school grades have improved over time but without any linked rise in student academic achievement measured by independent measures of attainment (ACT Assessment). This has been confirmed in several studies, sometimes leading to a 'Lake Wobegon effect' where everyone is declared 'above average' in attainment (Zirkel 1999)!

Norm-referencing had, by definition, previously worked to maintain results at a relatively constant level (Foxman 1997). Since grades were allocated proportionately the system of assessment would be unable to detect improvement in performance over time. Indeed the whole system was based on the assumption that change did not take place, and its *raison d'être* was therefore simply to discriminate between the performance of students in one cohort at a time. Student performance may have been improving prior to 1988. We have no way of judging. Since 1988 there is clear evidence of change over time. Not only is the entry cohort for examinations increasing as a proportion of the age cohort, the grades awarded to them are improving year-on-year. Nevertheless, because of concurrent changes, we are unable to attribute these improvements to market forces. A more

sophisticated comparison is required to separate the possible impact of market forces from changes in the nature of assessment.

A control group?

Perhaps what we need is the equivalent of a quasi-experimental 'control group'. One way of assessing the relative performance of state-funded schools before and after parental choice (even where high-quality 'before' figures are not available) is to use the fee-paying school sector as such a control group. On one interpretation, legislation such as the Education Reform Act 1988 has had no direct effect on fee-paying schools to whom it does not apply. Fee-paying schools are not even bound by the National Curriculum, although the majority follow an approximation of it since it was, in fact, designed on the basis of the traditional fee-paying curriculum. Such schools are and always have been in real market for customers - a market in which money changes hands, marketing is a significant activity, and schools open and close regularly (see Gorard 1997). We therefore have the basis of a (retrospective) natural experiment (see Bernard 2000, Gorard 2001). The treatment group consists of state-funded schools, and the control group is the fee-paying sector. The treatment is the introduction of the limited market which affected only the first group, whereas changes in the nature of assessment affected both groups. Possible confounds to this natural experiment are any changes in the type and proportion of fee-paying users over the period in question.

In general, the examination outcomes of students from fee-paying schools are superior to those from state-funded schools. Fee-paying schools regularly appear near the top of 'league tables' of raw-score results. What has changed since 1988? According to the figures in Table I, quite a lot has changed. As in the previous analysis, these figures are not ideal. In an ideal world they would be contextualised by changes in the gendered nature of fee-paying provision, the relative size of the sectors, differential examination entry policies, the impact of the Assisted Places Scheme, and other factors. Nevertheless, although earlier figures are not available by sector, it is clear that from 1992 onwards state-funded schools have been catching up with fee-paying schools at several levels of attainment. According to Howson (2000) this trend still continues. In fact, scores for fee-paying schools appear to have stalled (near the 100% barrier, see above), so that scores for the other sectors are catching up for as long as they improve year-on-year.

Both fee-paying and state-funded sectors have improved their scores for all three indicators presented here (and as shown in the previous section). These years represent the first in which cohorts recruited since 1988 were leaving school. They suggest that the basket of reforms within the state sector *did* lead to an improvement relative to the fee-paying sector, as well as an absolute increase in scores. Improvement ratios (see Gorard 2000c, 2001) show that state schools increased their qualification rate by 7% (81:76) while fee-paying schools increased theirs by 3% (89:86). State schools increased their GCSE benchmark by 26% (43:34) while fee-paying schools increased theirs by 19% (83:70). State schools increased their A levels scores by 19% (16.0:13.4) while fee-paying schools increased theirs by 15% (19.9:17.3).

Table I - Comparison of results by sector

	% 1GCSE A*-G		% 5GCSE A*-C		A levels points	
	LEA/GM	Fee	LEA/GM	Fee	LEA/GM	Fee
1992	76	86	34	70	13.4	17.3
1993	77	86	36	73	13.6	17.7
1994	80	87	39	75	14.5	20.0
1995	80	90	40	80	14.9	19.2
1996	80	91	41	79	15.5	19.3
1997	81	89	43	83	16.0	19.9

As with any 'experiment' it is important to replicate the results if possible. We are working on a characterisation of local areas where market activity is high, and where such activity is low. This dichotomy could be produced using local figures on population density, distance travelled, level of appeals against school allocation, distribution of surplus places, or school diversity. As above, the test would then be whether the gain score in terms of examination results was greater in the high activity area (compare the logic, but not the method, used in Levacic et al. 1998).

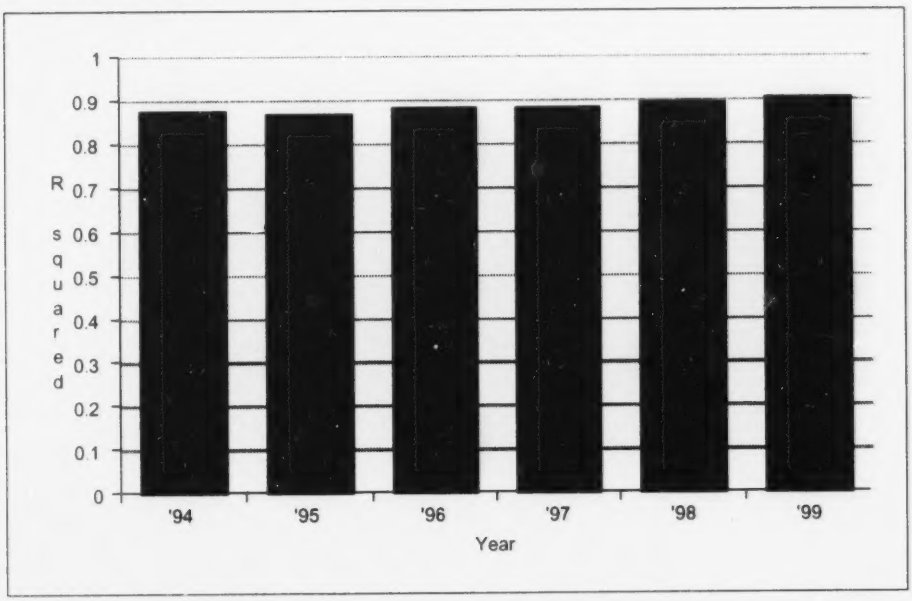
Regression models

A third way of assessing changes over time is based on the changing relationship between background variables (socio-economic context) and school attainment (outcome scores). For example it is clear that measures of student poverty such as eligibility for free school meals (FSM) and student achievement are strongly negatively related. This relationship holds at any level of aggregation from individual to national (e.g. Gorard 2000d). Over the last twelve years since the ERA 1988 figures for FSM have increased proportionately among the school population, and outcome scores such as GCSE benchmark figures have also increased. If both figures are increasing but are negatively related, then their precise relationship must be changing over time. One reasonable interpretation of a genuine improvement in an era of increasing raw-scores would be that outcomes are no longer as socially-determined as they were previously. Children from poor families would now be more likely to obtain their 'fair share' of the qualification spoils than they were in previous cohorts. Is this true?

Figure 3 suggests that this is not, in fact, so. If anything, the link between the explanatory variables such as poverty and outcomes scores such as GCSE results is growing slightly stronger (although this is likely to be an artefact of ongoing changes in the quality of figures in our historical database, see Gorard and Fitz 2000a). The R squared values represented here are derived from a series of regression analyses - one for each year for the 1,000 or so schools in the 40 LEAs selected as a sub-sample of our national database for more detailed study (the full list can be seen in White et al. 2001). Each analysis used the standard GCSE benchmark for each secondary school in England as the dependent variable, and various measures of school socio-economic composition as the independent variables (see Appendix). The independent variables were entered into the model using a forward stepwise approach.

Whatever the improvements in raw-sores over time (see above) it is clear that these have not 'broken' the well-established link between student background and school outcomes. In apparent confirmation of this, a study at the Centre for Longitudinal Studies in London has been reported as showing that children from poor families are no more likely to get qualifications than they were 20 years ago (Hackett 2000). Similar conclusions, but for different reasons and using different methods, have also been drawn in France (Duru-Bellat and Kieffer 2000).

Figure 3 - Variance explained by background factors over time



In every year the proportion of children defined as in poverty (taking or eligible for FSM) is a key predictor of school outcomes, as are the proportion of boys to girls, the existence of a sixth form, and the type of school (Community, Foundation, Grammar, Voluntary-aided RC, Voluntary-aided other, or Technology). The ethnic breakdown of the students was a significant predictor in 1994, but is no longer so. Rather the proportion of students with English as a second language has taken its place. Other factors that appear significant for one or more years, without showing any clear trend, are levels of unauthorised absence, size of school, and the proportion of students with special (additional) educational needs.

It should be recalled here that the values in Figure 3 are for R *squared*. The multiple correlations are higher again. It should also be noted that unlike standard 'value-added' models this analysis uses no scores for prior attainment (overcoming a major criticism of school effectiveness work such as Yang et al. 1999 and Goldstein et al. 2000 that it omits social factors, according to Rassool and Morley 2000). Around 90% of the variation in school outcomes can be explained by student background

characteristics and the nature of their school, and this figure is relatively constant over time (the independent variables for 1988/99 are more accurate than for other years). Given that these models also include an error component, there is little variance (from 100%) left to attribute to a school, or even a school system, effect. The possibility of discovering an improvement in this relatively small school effect over time would seem difficult enough. To partition out any of this improvement which is a direct result of market forces would appear nearly impossible.

In this we agree with the conclusion of Plewis (1999) that the most effective way to tackle inequality in education is by addressing poverty. The variation between school outcomes is very small (much smaller than within schools), so that strategies like the market which are aimed at schools or larger units like EAZs, rather than individuals, are likely to fail. 'Over the past 25 years... studies show that individual and family background traits explain the vast majority of the variance in student test scores, and observable school characteristics, such as per-pupil spending, teacher experience, or teacher degree level, have at best a weak relationship with student outcomes' (Goldhaber et al. 1999, p.199). Nevertheless, we shall continue with a more detailed school-level analysis, using more schools, more indicators, alternative models, and most importantly tracing the performance of schools back before 1994.

Differential attainment

As the relationship between student intake and school outcomes remains relatively stable over time while school outcomes scores have improved, it is little surprise to find that differences in attainment between identifiable social groups are declining. We have dealt with this decline in more detail elsewhere (e.g. Gorard et al. 1999b). Using valid proportionate analyses, differences in attainment have declined as measured between: the highest and lowest achievers; ethnic groups; boys and girls; economic regions, and school sectors. Despite the continued importance of socio-economic, as opposed to educational, determinants of school outcomes the system as a whole is therefore becoming fairer. This change is partly due to changes in the nature of the assessment system, and partly due to the changing patterns of socio-economic segregation between schools (Gorard and Fitz 2000b). There is little evidence as yet that this welcome, but limited, change is also related to the Education Reform Act 1988.

Discussion

Knowledge is not a static commodity, and comparisons of changes over time in school attainment have to try and take these changes into account. The complaint by the National Commission on Education (1993) that number skills have deteriorated over time for 11-15 year olds would have an analogy in the clear drop over the last millennium in archery standards among the general population. Nuttall (1979) used the example of the word 'mannequin' to make the same point. If the number of children knowing the meaning of this word drops from 1950s to the 1970s is this evidence of some kind of decline in schooling? Perhaps it is simply evidence that words and number skills have changed in their everyday relevance. On the other hand, if the items in any test are changed to reflect these changes in society, then how do we know that the test is of the same level of difficulty as its predecessor? In public examinations, by and large, we have until now relied on norm-referencing. That is, two tests are declared equivalent in difficulty if the same proportion of matched candidates obtain each graded result on both tests. The assumption is made that actual standards of each annual cohort are equivalent, and it is these that are used to benchmark the assessment. How then can we measure changes in standards over time? But, if the test is not norm-referenced how can we tell that apparent changes over time are not simply evidence of differentially demanding tests?

It has been claimed that the level of attainment required to gain Level 4 at KS2 has fallen over time. The evidence for this is that whereas students needed 52% to gain Level 4 English in 1997, the corresponding figures for 1998 and 1999 were 51% and 47% (Cassidy 1999). The response from the Qualifications and Curriculum Authority is that percentages are bound to change over time as the difficulty of the tests vary year-on-year, but that these differences are not educationally significant. A counter response has been that the QCA deliberately reduced the threshold because David Blunkett (Secretary of State for Education) had staked his career on 80% of 11 year-olds gaining Level 4 by 2002. Since in 1998 only 65% of the population gained Level 4, it is claimed that while the target has been retained the pass-mark has been conveniently lowered. An independent enquiry was ordered, the results of which have mainly supported the QCA position. This debate encapsulates the problems of discussing changes in assessments over time.

When serious attempts have been made to compare standards of attainment over time, and taking into account all of the above caveats, the results are generally that standards are *not* falling. In some cases there is no firm evidence of change, and in others there are improvements over time. For example, an analysis of successive GCSE cohorts from 1994 to 1996 found a significant improvement in performance over time (Schagen and Morrison 1998). It is possible to question the reality of this improvement in strict criterion-referenced terms, but there is at any rate no evidence of any decline, and some suggestion that things are actually getting better.

The most obvious conclusion to be drawn from this consideration is that there is no easy answer to the question 'have standards improved as a result of market forces in education?'. Even our relatively simple prior question about the impact of market forces on school compositions is still the subject of much debate (e.g. Gorard 2000e). In that ongoing debate over composition, the opponents of school choice predicted a rise in socio-economic segregation between schools which has never been demonstrated in fact. A similar situation applies here. Advocates of school choice predicted a rise in standards. We have not been able to demonstrate successfully that this has occurred, except in relation to fee-paying schools. Standards have clearly not declined since 1998 in England and Wales. Whether they have improved and, if so, whether this improvement is attributable to market forces is still unclear. Opponents of choice may say that this is because choice does not work to drive up school standards. Advocates of choice may argue that the methodological difficulties involved now make their thesis untestable. A more neutral observer might point out the very limited nature of the market in schools anyway. It would however be ironic if the long-term impact of choice was found to lead to no difference in standards but an amelioration in school segregation. This would be precisely contrary to the views of choice advocates (wishing to drive up standards) and of their opponents (fearing social polarisation).

Acknowledgements

The authors would like to thank John Fitz and Patrick White. This work was funded by the ESRC (grant number R000238031).

References

Baird, J., Cresswell, M., Newton, P. (2000) Would the real gold standard please step forward?, *Research Papers in Education*, 15, 2, 213-229

Bernard, H. (2000) *Social research methods: qualitative and quantitative approaches*, London: Sage

Cassidy, S. (1999) Pass mark 'fiddle' is strenuously denied, *Times Educational Supplement*, 28/5/99, p. 2

Cassidy, S, Mansell, W. and Hackett, G. (2000) Test marks queried by one in 7 primaries, *Times Educational Supplement*, 23/6/00, p.1

Chubb, J. and Moe, T. (1990) *Politics, markets and America's schools*, Washington: Brookings Institute

Coleman, J., Hoffer, T. and Kilgore, S. (1982) *High School Achievement: Public, private and Catholic schools compared*, New York: Basic Books

Cookson, P. (1994) *School choice*, Yale: Yale University Press

DfEE (1998) *Statistics of Education: Public Examinations GCSE/GNVQ and GCE AGNVQ in England 1998*, Government Statistical Service

Dobson, J. and Henthorne, K. (1999) Leavers who don't fare well, *Times Educational Supplement*, 29/10/99, p.19

Duru-Bellat, M. and Kieffer, A. (2000) Inequalities in educational opportunities in France: educational expansion, democratization or shifting barriers?, *Journal of Education Policy*, 15, 3, 333-352

Erickson, D. (1989) A libertarian perspective on schooling, in Boyd, W. and Cibulka, J. (Eds.) *Private schools and public policy*, London: Falmer Press

Foxman, D. (1997) *Educational league tables: for promotion or relegation?*, London: Association of Teachers and Lecturers

Friedman, M. and Friedman, R. (1980) *Free to choose*, Harmondsworth: Penguin

Fuller, B., Elmore, R. and Orfield, G. (1996) Policy-making in the dark, in Fuller, B. and Elmore, R. (Eds.) *Who chooses? Who loses?*, New York: Teachers College Press

Goldhaber, D., Brewer, D. and Anderson, D. (1999) A three-way error component analysis of educational productivity, *Education Economics*, 7, 3, 199-208

Goldstein, H., Huiqi, P., Rath, T. and Hill, N. (2000) *The use of value-added information in judging school performance*, London: Institute of Education

Gorard, S. (1997) *School Choice in an Established Market*, Aldershot: Ashgate

Gorard, S. (1999a) 'Well. That about wraps it up for school choice research': a state of the art review, *School Leadership and Management*, 19, 18, 25-47

Gorard, S. (1999b) Examining the paradox of achievement gaps, *Social Research Update*, 26, 1-4

Gorard, S. (1999c) Keeping a sense of proportion: the "politician's error" in analysing school outcomes, *British Journal of Educational Studies* , 47, 3, 235-246

Gorard, S. (2000a) *Education and Social Justice*, Cardiff: University of Wales Press

Gorard, S. (2000b) One of us cannot be wrong: the paradox of achievement gaps, *British Journal of Sociology of Education*, 21, 3, 391-400

Gorard, S. (2000c) Questioning the crisis account: a review of evidence for increasing polarisation in schools, *Educational Research*, 42, 3, 309-321

Gorard, S. (2000d) 'Underachievement' is still an ugly word: reconsidering the relative effectiveness of schools in England and Wales, *Journal of Education Policy* , 15, 5, 559-573

Gorard, S. (2000e) Here we go again: a reply to 'What's in a number?' by Gibson and Asthana, *Research Papers in Education* , 15, 2, 155-162

Gorard, S. (2001) *Quantitative Methods in Educational Research: The role of numbers made easy*, London: Continuum

Gorard, S. and Fitz, J. (2000a) Investigating the determinants of segregation between schools, *Research Papers in Education* , 15, 2, 115-132

Gorard, S. and Fitz, J. (2000b) Markets and stratification: A view from England and Wales, *Educational Policy* , 14, 3, 405-428

Gorard, S., Rees, G. and Selwyn, N. (1999a) *Lifetime learning targets in Wales: a research summary*, Cardiff: National Assembly for Wales, 81 pages

Gorard, S., Salisbury, J. and Rees, G. (1999b) Reappraising the apparent underachievement of boys at school, *Gender and Education*, 11, 4, 441-454

Hackett, G. (2000) Children still class bound, *Times Educational Supplement*, 21/4/00, p.11

Hackett, G. and Kelly, A. (2000) Heads question literacy gains, *Times Educational Supplement*, 14/7/00, p. 1

Haque, Z. and Bell, J. (2000) *Evaluating the performances of minority ethnic pupils in secondary schools*, presentation at BERA September 2000, Cardiff

Hardman, J. and Levacic, R. (1997) The impact of competition on secondary schools, in Glatter, R., Wood, P. and Bagley, C. (Eds.) *Choice & Diversity in schooling: perspectives and prospects*, London: Routledge

Howson, J. (1999) Numbers of special schools still falling, *Times Educational Supplement*, 29/10/99, p.20

Howson, J. (2000) Solid state, *Times Educational Supplement*, 14/7/00, p. 24

Jeynes, W. (2000) Assessing school choice: a balanced perspective, *Cambridge Journal of Education*, 30, 2, 223-241

Levacic, R., Hardman, J. and Woods, P. (1998) *Competition as a spur to improvement? Differential improvement in GCSE examination results*, presentation to International Congress for School Effectiveness and Improvement, Manchester

Mansell, W. (2000) Missing pupils skew exam data, *Times Educational Supplement*, 23/6/00, p.3

McGill, P. (2000) Unfair grades for 11-plus pupils, *Times Educational Supplement*, 14/4/00, p.7

National Commission on Education (1993) *Learning to succeed*, London: Heinemann

Noah, H. and Eckstein, M. (1992) Comparing secondary school leaving examinations, in Eckstein, M. and Noah, H. (Eds.) *Examinations: comparative and international studies*, Oxford: Pergamon Press

Nuttall, D. (1979) The myth of comparability, *Journal of the National Association of Inspectors and Advisers*, 11, 16-18

Plewis, I. (1999) Educational inequalities and education action zones, in Pantazis, C. and Gordon, D. (Eds.) *Tackling inequalities*, Bristol: Policy Press

Powers, J. and Cookson, P. (1999) The politics of school choice research: fact fiction and statistics, *Educational Policy*, 13, 1, 104-122

Rassool, N. and Morley, L. (2000) School effectiveness and the displacement of equity discourses in education, *Race Ethnicity and Education*, 3, 3, 237-258

Robinson, P. (1997) *Literacy, numeracy and economic performance*, London School of Economics: Centre for Economic Performance

Schagen, I. and Morrison, J. (1998) *QUASE Quantitative analysis for self-evaluation: Overview report 1997: Analysis of GCSE cohorts 1994 to 1996*, Slough: NfER

Strand, S. (2000) *Pupil mobility, attainment and progress during Key Stage 1: a case study in caution*, presentation at BERA September 2000, Cardiff

TES (1999) Raising standards faster, Times *Educational Supplement*, 12/11/99, p. 7

West, A., Pennell, H. and West, R. (2000) New Labour and school-based education in England: changing the system of funding?, *British Educational Research Journal*, 26, 4, 523-536

White, P., Gorard, S., Fitz, J. and Taylor, C. (2001) Regional and local differences in admission arrangements for schools, *Oxford Review of Education*, 27, 3

Willms, J. (1992) *Monitoring school performance. A guide for educators*, London: Falmer Press

Witte, J. (1990) Introduction, in Clune, W. and Witte, J. (Eds.) *Choice and control in American education, Volume 1*, London: Falmer Press

Yang, M. and Woodhouse, G. (2000) *Progress from GCSE to A and AS level: institutional and gender differences, and trends over time*, presentation at BERA September 2000, Cardiff

Yang, M., Goldstein, H., Rath, T. and Hill, N. (1999) The use of assessment data for school improvement purposes, *Oxford Review of Education*, 25, 4, 469-483

Zirkel, P. (1999) Grade inflation: a leadership opportunity for schools of education?, *Teachers College Record*, 101, 2, 247-260

Appendix

Nationally, using all secondary schools in England and Wales (as well as all primary schools, not used here), our database includes the following figures from school census returns. These are organised in order of the 15 orthogonal factors emerging from a principal components analysis of the figures for all years combined. Thus poverty, ethnicity and first language appear as one factor, absence from school and GCSE results appear as two factors (and so on).

1. Proportion of black or other students, taking/eligible for FSM, and speaking ESL
2. Unauthorised absence, proportion obtaining 5 GCSE A-G, and 1 GCSE A-G
3. Authorised absence, proportion obtaining 5 GCSE A-C
4. Number of pupils, proportion of white students
5. Existence of sixth form
6. Proportion of students with SEN statements
7. Proportion of Asian students
8. Proportion of students with unstatemented SEN
9. Girls only, grammar school
10. Foundation school
11. Boys school
12. 13-18 school
13. VA other school
14. VA RC school
15. Technology school